MULTISCALE COLOR-TEXTURE IMAGE SEGMENTATION WITH ADAPTIVE REGION MERGING

Tao Wan   Nishan Canagarajah   Alin Achim

Department of Electrical and Electronic Engineering
University of Bristol, Bristol, BS8 1UB, UK
{Tao.Wan, Nishan.Canagarajah, Alin.Achim}@bristol.ac.uk

ABSTRACT

A novel multiscale image segmentation algorithm is presented, which is based on the dominant color and homogeneous texture features (HTF) that are adopted in the MPEG-7 standard. These features are efficiently combined to perform the automatic segmentation. First, the image is roughly segmented into textured and nontextured regions using Gabor decomposition. A multiscale segmentation is then applied to the resulting regions, according to the local texture feature. Finally, a precise boundary refinement procedure is employed to accurately determine the boundaries between textured and nontextured regions. A novel region merging algorithm is introduced with a simple and effective segment classification by using HTF to deal with the over-segmentation problem. Experiments show that our algorithm provides an improved performance compared with JSEG and a watershed algorithm.

Index Terms— multiscale segmentation, dominant color, homogeneous texture, adaptive region merging, MPEG-7

1. INTRODUCTION

Image segmentation has become one of the most popular and challenging problems in image processing. Recently, significant progress has been made in texture segmentation [1][2] and color segmentation [3][4]. However, the area of combined color and texture segmentation is relatively new and still has open problems that need to be further investigated. If an image contains only homogeneous color regions, direct clustering methods in color space are sufficient to handle the problem. In reality, natural scenes are rich of nonuniform color and texture due to effects of lighting, perspective, shadow etc. In considering these factors, we propose a novel approach that focuses on combining the dominant color [5] and the homogeneous texture feature [6] based on the MPEG-7 standard descriptors for visual contents.

The dominant color provides a compact description of the representative colors in an image or an image region. The homogeneous texture feature characterizes the region texture using the mean energy and the energy deviation from Gabor subband features. These two content descriptors have been demonstrated to be effective for the color and texture features analysis in image retrieval applications [5][7]. In this paper, dominant color and homogeneous texture features are integrated into a single image segmentation framework.

A multiscale segmentation is applied to textured and nontextured image regions, which are crudely segmented using Gabor-filtered features. Hence, the image is initially partitioned into relatively homogenous regions but with rough boundaries between textured and nontextured regions. Then boundary refinement is employed to delineate these boundaries more accurately. A novel region merging approach classifies image segments into smooth, textured and complex categories. Color and texture similarity distances are measured on smooth and textured classes respectively, or combined into a complex class. The key to the success of the proposed algorithm is that the entire image can be treated as a set of individual color or textured patches depending on the local image characteristics. Experiments show that the proposed algorithm outperforms state-of-the-art segmentation algorithms, such as JSEG [4] and the watershed algorithm [2]. The former applies the same scale to the whole image without taking into account the local region difference, while the later generates combined gradient images, which is computationally more complex.

The structure of the paper is as follows: the texture segmentation using Gabor filters is described in Section 2. The multiscale image segmentation algorithm is discussed in Section 3. Our proposed adaptive region merging approach and experimental results are presented in Section 4 and Section 5 respectively. Finally, conclusions are drawn in Section 6.

2. TEXTURE SEGMENTATION

Gabor filters have been widely used to extract texture features. This is motivated by the fact that a set of ideal filter banks can be considered as orientation and scale tunable edge and

Tao Wan was supported by an Overseas Research Students Award (ORSAS).

(a) (b) (c)

Fig. 1. Texture segmentation (a) Original image. (b) Texture map using Gabor decomposition. (c) Texture map using Steerable decomposition. White regions correspond textured regions, and black regions correspond nontextured regions.

line detectors, and the statistics of these micro features can be used to characterize the underlying texture [7]. In this work, we use a five-scale Gabor filter decomposition with six orientations addressed in [6]. Following the arguments in [8], one can expect to obtain better results by using a finer scale and orientational decomposition, in a similar way to the model of the human visual cortex.

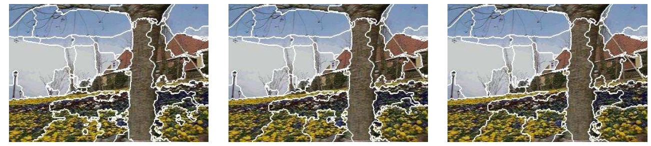
The goal of this stage is to obtain a crude segmentation map with only textured and nontextured regions. The specific value of Gabor decomposition at each pixel location doesn't provide much information. Therefore, the maximum operator is applied to obtain the feature value $T_{max}(x, y)$ at the pixel location (x, y). The additional advantage is that there is a reduction in computational complexity.

$$T_{max}(x, y) = max\{g_i(x, y)\} \quad i = 1, 2, 3, \ldots, 30 \tag{1}$$

where g_i is the i^{th} subband Gabor coefficients. In order to obtain a uniform characterization of texture, median filtering is employed on $T_{max}(x, y)$ to filter out the texture associated with transition between regions. Finally, a two-level K-means algorithm is used to segment the image into textured and nontextured regions. A similar method was presented in [9], in which Chen *et al.* used a one-level steerable filter decomposition with four orientations. Fig.1 shows the segmentation results obtained using both algorithms. Unlike the method in [9], which focuses on low-resolution compressed images, our algorithm could handle any type of natural images.

3. MULTISCALE IMAGE SEGMENTATION

The textured and nontextured regions are further segmented into relatively small and homogeneous regions, while retaining the boundaries between the two regions. The dominant colors are first extracted based on Peer Group Filtering (PGF) [10] and the Generalized Lloyd Algorithm [11]. Then, the JSEG algorithm proposed by Deng *et al.* in [4] is used to minimize the cost associated with partitioning an image at different scales. A bigger window size is used for high scales, which are useful for detecting texture boundaries, while lower scales are employed in order to localize the intensity of color edges. It is reasonable to apply the lower scales to the nontextured region, which has a more or less homogeneous texture, while higher scales are adopted for the textured region to find the texture boundaries. In contrast with JSEG, which doesn't take into account the local texture difference between the image regions, the strength of this approach is that we are able to apply the multiscale segmentation simultaneously to the same image according to the local texture characteristics.



(a) (b) (c)

Fig. 2. Boundary refinement results (a) $\alpha = 0.8, \beta = 0.0$. (b) $\alpha = 0.0, \beta = 0.8$. (c) $\alpha = 1.0, \beta = 0.8$.

However, the current boundary locations between textured and nontextured regions are not the actual boundaries due to the fact that K-means clustering can only segment the image into rough regions. Moveover, multiscale segmentation provides accurate results only within the textured and nontextured regions. Consequently, a boundary refinement step is employed to adjust the boundaries between the two regions. A pixel is assigned to the neighbor class that has the minimum D value using the following function:

$$D = Dist(C^0, C^i) + \alpha(S_4^i - D_4^i) + \beta(S_8^i - D_8^i) \tag{2}$$

Where $Dist$ refers to the Euclidean distance measure, C^0 and C^i are the dominant color vectors of the current pixel and its i^{th} neighbor segment, S_4^i and S_8^i are the numbers of 4- and 8-neighbor pixels belonging to the ith segment, while D_4^i and D_8^i are the numbers of 4- and 8-neighbor pixels belonging to the different classes of the i^{th} segment. α and β represent the strength of the spatial constraint. Specifically, as α and β increase, a pixel is more likely to belong to the class to which many of its neighbors belong. Thus region boundary smoothness is achieved. The influence of α and β on the boundary refinement procedure is shown in Fig.2. The result in Fig.2(c) that uses higher values of α and β has smoother boundaries compared with Fig.2(a) and 2(b).

4. ADAPTIVE REGION MERGING

In general, the result of applying the algorithm described in the previous sections leads to over-segmentation. In this work, segmented regions are therefore considered individually rather than globally. The segments are classified into three categories. The segments with more than 80% of their pixels belonging to the nontextured area are categorized as smooth segments. Similarly, the segments with more than 80% of their

(a) (b)

Fig. 3. Adaptive region merging results. (a) Image segmentation based on three-category. (b) Image segmentation based on two-category.

pixels in the textured area are classified as textured segments. The remaining segments are classed as complex, which means that they do not have a dominant feature. It is worth noting that in order to implement this stage of the algorithm only the coefficients of Gabor-filtered outputs are used.

A corresponding merging criterion is provided for each category. The main difference lies in the way of calculating the feature distance. Smooth segments are merged based on their dominant color similarity. To achieve this, the Euclidean distance of the color histograms extracted from the neighboring smooth segments is calculated. For textured segments, region similarity is computed using energies $[e_1, \ldots, e_{30}]$ and energy deviations $[d_1, \ldots, d_{30}]$ of Gabor-filtered coefficients as suggested in [6]. The Euclidean distance of HTF is measured between the neighboring textured segments. The HTF extraction methodology has been proposed in [6] for the purpose of image retrieval. Here it is applied locally to image regions, rather than globally to the whole image.

The merging criterion for a complex segment is more difficult than for the other two categories as these segments contain both color and texture features. First, we merge the image into homogenous textured segments by using the textured-segment method. Then, the smooth-segment merging is applied. This approach can be extended to use a different combination of color and texture features. The pair of regions with the minimum distance is merged. The process continues until a maximum threshold of the distance is reached. It has been found experimentally that the best value of the threshold is in the range of 0.4 to 0.6.

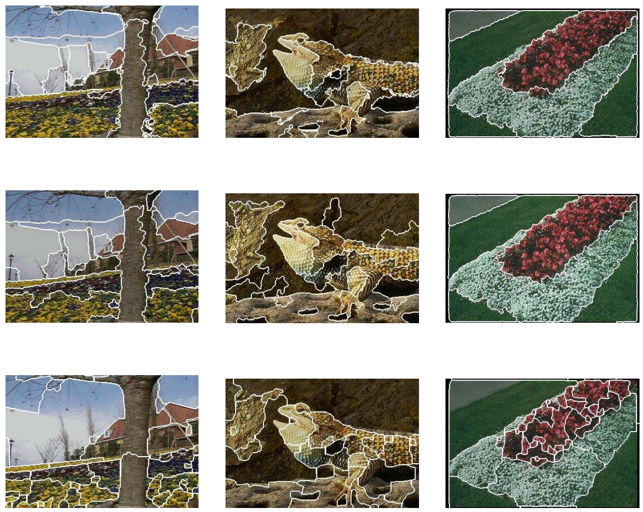
A two-category approach is also investigated, where only smooth and textured segments are considered. The merging criteria described above for the smooth and textured segments are used. The three-category method performs better than the two-category method as shown in Fig.3. The main reason is that it is impossible to separate the color and spatial frequency components of texture completely. Adaptive region merging provides a better way of understanding image features.

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

The segmentation algorithm is tested on a variety of natural images. Fig.4 shows the segmentation results, comparing with the segmentation obtained by JSEG [4] and the watershed algorithm [2]. The JSEG examples shown are set to the default options which are determined automatically. The watershed method applies the region-depth threshold to the gradient surface, which is set as 0.15 times the gradient, then followed by a spectral clustering approach to group together these initial regions. There are no distinct differences between our algorithm and JSEG in Fig.4(c). However, in the watershed segmentation, the grass land is merged with the road on the top left corner of the image, and the flower land is also over-segmented. Significant differences can be seen in Fig.4(a), in that the road lamps are well segmented by our algorithm, but half merged with one of the sky segments by JSEG, and have the wrong boundary in the watershed. Similarly, the top branch of the tree is segmented well, while in JSEG it is merged into the sky. In the watershed algorithm, the main trunk of the tree is not well segmented from the background. The flower lands are merged into one segment on the left side of the tree in our algorithm due to HTF similarity in spite of their color features being different. In Fig.4(b), the rocks around the lizard and the skin area of the lizard have much clearer boundaries than in JSEG and the watershed. Moreover, as shown in Fig.5, our algorithm outperforms JSEG in finding the salient regions of the color images. The hand-labeled ground truth segmentation images from Berkeley dataset [12] are also illustrated in Fig.5. In Fig.5(a), a leopard in the foreground, can also be a salient region shown in the ground truth image. The proposed algorithm segments the contour of the leopard better than JSEG does. Similarly, in Fig.5(b) and 5(c), the starfish and the tiger, which are salient objects in both images respectively, are better segmented than by JSEG. The watershed is also able to accurately detect contours but still exhibits an over-segmentation problem within salient objects.

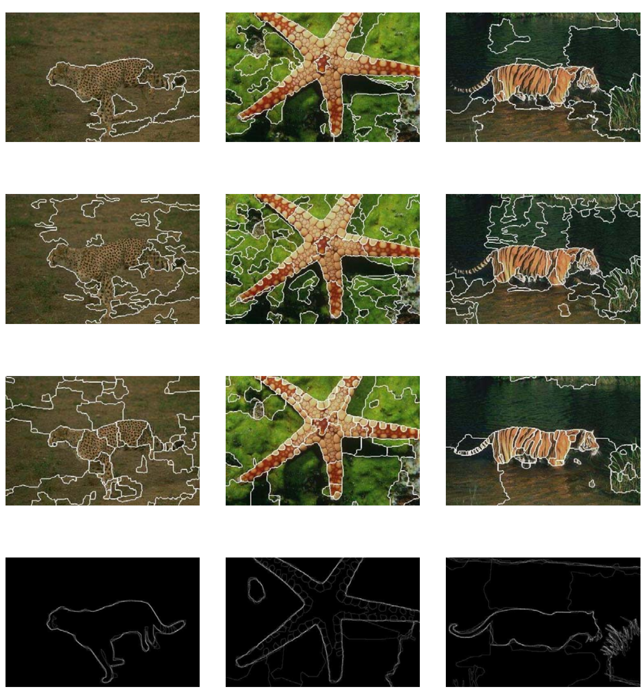
6. CONCLUSIONS AND FUTURE WORK

A new image segmentation algorithm was presented, which is based on low-level features of color and texture. The performance of the proposed algorithm was assessed on a variety of still images of natural scenes. The main contribution of this work is that it provides a robust and accurate automatic image segmentation and highlights salient objects within the image. This is important for the semantic understanding of images. Future work will consider several improvements to the method reported here. These will include the implementation of a better clustering algorithm, as well as an improved similarity distance calculation approach. The former will help in obtaining more accurate initial texture segmentation results, while the later could enhance the adaptive region merging step



(a) (b) (c)

Fig. 4. Segmentation results. From top to bottom, results based on multiscale segmentation and adaptive region merging, results using JSEG [4], and results using the watershed algorithm [2]



(a) (b) (c)

Fig. 5. Segmentation results. From top to bottom, results based on multiscale segmentation and adaptive region merging, results using JSEG [4], results using the watershed algorithm [2], and hand-labeled ground truth image segmentation from Berkeley dataset [12]

of our proposed algorithm.

7. REFERENCES

[1] P. R. Hill, C. N. Canagarajah, and D. R. Bull, "Image segmentation using a texture gradient based watershed transform," *IEEE Tran. Image process.*, vol. 12, no. 12, Dec. 2003.

[2] R. J. O'Callaghan and D. R. Bull, "Combined morphological-spectral unsupervised image segmentation," *IEEE Tran. Image process.*, vol. 14, no. 1, Jan. 2005.

[3] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Tran. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, Jan. 2004.

[4] Y. Deng, B. S. Manjunath, and H. Shin, "Color image segmentation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1999.

[5] Y. Deng, B. S. Manjunath, C. Kenney, M. S. Moore, and H. Shin, "An efficient color representation for image retrieval," *IEEE Tran. Image process.*, vol. 10, no. 1, Jan. 2001.

[6] Y. M. Ro, M. C. Kim, H. K. Kang, B. S. Manjunath, and J. W. Kim, "MPEG-7 homogeneous texture descriptor," *ETRI Journal*, vol. 23, no. 2, pp. 41–51, Jun. 2001.

[7] B. S. Manjunath, P. Wu, S. Newsam, and H. D. Shin, "A texture descriptor for browsing and similarity retrieval," *Journal of Signal Processing: Image Commun.*, vol. 16, no. 1-2, pp. 33–43, Sep. 2000.

[8] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biological Cybernetics*, vol. 61, pp. 103–113, 1989.

[9] J. Chen, T. N. Pappas, A. Mojsilovic, and B. Rogovitz, "Adaptive perceptual color-texture image segmentation," *IEEE Tran. Image process.*, vol. 14, no. 10, Oct. 2005.

[10] Y. Deng, C. Kenney, M. S. Moore, and B. S. Manjunath, "Peer group filtering and perceptual color image quantization," in *Proc. IEEE Int. Symp. Circuits and Systems VLSI*, Orlando, FL, Jun. 1999, vol. 4, pp. 21–24.

[11] A. Gersho and R. M. Gray, *Vector quantization and signal compression*, Kluwer Academic Publisher, Norwell, Mss., 1993.

[12] *The Berkeley Segmentation Dataset Benchmark*, University of California, Berkeley, CA, http://www.eecs.berkeley.edu/research/projects/cs/vision/grouping/segbench/ edition, July 2003.